

# Some Asymptotic Results for Fiducial and Confidence Distributions

Piero Veronese <sup>1</sup>and Eugenio Melilli

Bocconi University, via Röntgen, 1, 20136, Milan, Italy

piero.veronese@unibocconi.it, eugenio.melilli@unibocconi.it

**Abstract.** Under standard regularity assumptions, we provide simple approximations for fiducial and confidence distributions and discuss their connections with objective Bayesian posteriors. For a real parameter the approximations are accurate at least to order  $O(n^{-1})$ . For a multivariate parameter indexing an exponential family we show that the asymptotic fiducial distribution is normal and coincides with the asymptotic Bayesian posterior.

**Keywords:** ancillary statistic, confidence curve, natural exponential family, matching prior, reference prior.

## 1 Introduction

Confidence and fiducial distributions, often confused in the past, have recently received a renewed attention by statisticians thanks to some contributions which clarify the concepts within a purely frequentist setting and overcome the lack of rigor and completeness typical of the original formulations. The book by Schweder & Hjort (2016) gives a wide and comprehensive presentation of the theory of confidence distributions and includes a complete list of references on the topic. For what concerns fiducial distributions Hannig, starting from the original idea of Fisher, has developed in several papers a *generalized fiducial inference* which is suitable for a large range of situations; see Hannig et al. (2016) for a complete review.

Given a random vector  $\mathbf{S}$  (representing the observations or a sufficient statistic summarizing them) and a real parameter of interest  $\theta$  indexing the distribution function  $F_\theta$  of  $\mathbf{S}$ , a *confidence distribution* (CD) is a function  $C$  of  $\mathbf{S}$  and  $\theta$  such that: i)  $C(\mathbf{s}, \cdot)$  is a distribution function on  $\mathbb{R}$  for any fixed realization  $\mathbf{s}$  of  $\mathbf{S}$  and ii)  $C(\mathbf{S}, \theta)$  has a uniform distribution on  $(0, 1)$ , whatever the true value of  $\theta$ . The relevant condition is the second one which implies that the coverage of intervals derived from  $C$  is exact. If the second assumption is satisfied only for the sample size tending to infinity,  $C$  is an *asymptotic* CD and the coverage is correct only approximately. Given a CD, it is

---

<sup>1</sup>Corresponding author

possible to define the so-called *confidence curve*  $cc_{\mathbf{s}}(\theta) = |1 - 2C(\mathbf{s}, \theta)|$ , which displays the confidence intervals induced by  $C$  for all levels, see for example Figure 2.

In order to define a generalized fiducial distribution, Hannig starts from a *data-generating equation*  $\mathbf{S} = \mathbf{G}(\mathbf{U}, \boldsymbol{\theta})$ , where  $\mathbf{U}$  is a random vector with known distribution, and use it to transfer randomness from  $\mathbf{S}$  to  $\boldsymbol{\theta}$ . For real  $\theta$  and  $S$ , a natural choice of  $G(U, \theta)$  is  $F_{\theta}^{-1}(U)$  which leads to a fiducial distribution (FD) equal to that originally proposed by Fisher (1930), namely

$$h_s(\theta) = \left| \frac{\partial}{\partial \theta} F_{\theta}(s) \right|. \quad (1)$$

In this paper we provide some asymptotic results for FDs and CDs under standard regularity conditions and highlight the connections with objective Bayesian inference which, as well known, produces posterior distributions free of any subjective prior information. After some preliminaries (Section 2), we provide the first-order asymptotic expansion for an FD and/or a CD of a real parameter and show that it coincides with that of the Jeffreys posterior (Section 3.1). An approximation based on the  $p^*$ -formula of Barndorff-Nielsen (1980, 1983) is given in Section 3.2. Finally, in Section 4 we show that the asymptotic FD for a multivariate parameter of an exponential family is normal and that it coincides with the corresponding asymptotic Bayesian posterior.

## 2 Preliminaries

Consider first an i.i.d. sample from a real regular natural exponential family (NEF) and let  $S$  be the corresponding sample sum.  $S$  is a sufficient statistic with density, with respect to a suitable measure,  $p_{\theta}(s) = \exp\{\theta s - M(\theta)\}$ ,  $s \in \mathbb{R}$ ,  $\theta \in \Theta \subseteq \mathbb{R}$  and distribution function  $F_{\theta}(s)$ . Veronese & Melilli (2015) show that

$$H_s(\theta) = 1 - F_{\theta}(s) = \Pr_{\theta}(S > s) \quad (2)$$

is both an FD and a CD (possibly asymptotic) with density equal to (1). When  $F_{\theta}$  is discrete, a *left fiducial* distribution  $H_s^{\ell}$  can also be defined using  $F_{\theta}^{\ell}(s) = \Pr_{\theta}(S < s)$  instead of  $F_{\theta}(s)$ . To overcome this non-uniqueness,  $H_s$  and  $H_s^{\ell}$  can be mixed in different ways. In particular, the distribution function  $H_s^G$ , whose density  $h_s^G$  is proportional to the geometric mean of  $h_s$  and  $h_s^{\ell}$ , namely  $h_s^G \propto \sqrt{h_s h_s^{\ell}}$ , has good properties and is preferable to the arithmetic mean  $H_s^A = (H_s + H_s^{\ell})/2$ , see Veronese & Melilli (2016).

If an FD coincides with a Bayesian posterior distribution, the prior inducing this latter is called *fiducial prior* by Veronese & Melilli (2015). They show, using  $H_s^G$  for discrete models, that such a prior exists and coincides with the Jeffreys prior only for binomial, Poisson, negative-binomial, normal (with known variance) and gamma (with known scale) models. This result provides a

connection between fiducial inference and objective Bayesian analysis and extends a statement given by Lindley (1958). When a fiducial prior does not exist, Veronese & Melilli (2015) show that, for both continuous and discrete (lattice) models,  $H_s$  and  $H_s^A$  present the same two-term Edgeworth expansion of the Jeffreys posterior distribution function,  $\Pi^J(\cdot|s)$  say. Here we show that this is true also for  $H_s^G$ .

**Proposition 1** *Consider an i.i.d. sample of size  $n$  from a discrete real NEF on a lattice, with mean parameter  $\mu$  and variance function  $V(\mu)$ . Denoting the sample mean by  $\bar{x}_n = s/n$ , the two-term Edgeworth expansion of the fiducial distribution function  $H_s^G(z)$  of  $Z = \sqrt{n}(\mu - \bar{x}_n)$  is*

$$H_s^G(z) = \Phi\left(\frac{z}{\sqrt{V(\bar{x}_n)}}\right) - \phi\left(\frac{z}{\sqrt{V(\bar{x}_n)}}\right) \frac{V'(\bar{x}_n)}{V(\bar{x}_n)^{3/2}} \frac{2z^2 + V(\bar{x}_n)}{6} n^{-1/2} + O(n^{-1}), \quad (3)$$

where  $\Phi$  and  $\phi$  are the distribution function and density of the standard normal, respectively. Moreover, the expansion (3) coincides with those of  $H_s^A(z)$  and  $\Pi^J(z|s)$ .

For a parameter  $\theta$  in  $\mathbb{R}^d$ , inspired by the step-by-step procedure proposed by Fisher (1973), Veronese & Melilli (2016) give a simple and quite general definition of FD which we summarize here. See the latter paper for details and applications concerning the remaining part of the section.

Given a random vector  $\mathbf{S}$ , representing the sample or a sufficient statistic, with dimension  $m \geq d$  and density  $p_\theta$ , consider the partition  $\mathbf{S} = (\mathbf{S}_{[d]}, \mathbf{S}_{-[d]})$ , where  $\mathbf{S}_{[d]} = (S_1, \dots, S_d)$  and  $\mathbf{S}_{-[d]} = (S_{d+1}, \dots, S_m)$ , and suppose that  $\mathbf{S}_{-[d]}$  is ancillary for  $\theta$ . Clearly, if  $d = m$ ,  $\mathbf{S}_{-[d]}$  disappears. Thus, the density  $p_\theta$  of  $\mathbf{S}$  can be written as  $p_\theta(\mathbf{s}_{[d]}|\mathbf{s}_{-[d]})p(\mathbf{s}_{-[d]})$  and the information on  $\theta$  provided by the whole sample is included in the conditional distribution of  $\mathbf{S}_{[d]}$  given  $\mathbf{S}_{-[d]}$ . Assume now that there exists a one-to-one smooth reparameterization from  $\theta$  to  $\phi = (\phi_1, \dots, \phi_d)$ , with the  $\phi_i$ 's ordered with respect to their inferential importance, such that

$$p_\phi(\mathbf{s}_{[d]}|\mathbf{s}_{-[d]}) = \prod_{k=1}^d p_{\phi_{d-k+1}}(s_k|\mathbf{s}_{[k-1]}, \mathbf{s}_{-[d]}; \phi_{[d-k]}), \quad (4)$$

with obvious meaning for  $\mathbf{s}_{[0]}$  and  $\phi_{[0]}$ . If for each  $k$ , the one-dimensional conditional distribution function of  $S_k$  is monotone and differentiable in  $\phi_k$  and has limits 0 and 1 when  $\phi_k$  tends to the boundaries of its parameter space (this is always true, for example, if this distribution belongs to a regular real NEF), it is possible to define the joint fiducial density of  $\phi$  as

$$h_\phi(\phi) = \prod_{k=1}^d h_{\mathbf{s}_{[k]}, \mathbf{s}_{-[d]}}(\phi_{d-k+1}|\phi_{[d-k]}), \quad (5)$$

where

$$h_{\mathbf{s}_{[k]}, \mathbf{s}_{-[d]}}(\phi_{d-k+1}|\phi_{[d-k]}) = \left| \frac{\partial}{\partial \phi_{d-k+1}} F_{\phi_{d-k+1}}(s_k|\mathbf{s}_{[k-1]}, \mathbf{s}_{-[d]}; \phi_{[d-k]}) \right|. \quad (6)$$

Some results useful in the sequel follow.

- i) If there exists a sufficient statistic with the same dimension of the parameter ( $m = d$ ), an ancillary statistic is not needed. A relevant case is  $m = d = 1$ , so that formulas (5) and (6) reduce to  $h_s(\phi) = |\partial F_\phi(s)/\partial \phi|$ , the original proposal of Fisher (1930).
- ii) If one is interested in  $\phi_1$  only, from (4) it follows that it is enough to consider

$$h_s(\phi_1) = \left| \frac{\partial}{\partial \phi_1} F_{\phi_1}(s_d | \mathbf{s}_{[d-1]}, \mathbf{s}_{-[d]}) \right|,$$

which does not lose any sample information because it depends on all observations. Furthermore, notice that  $h_s(\phi_1)$  is also a CD.

- iii) If  $(\mathbf{S}_{[k-1]}, \mathbf{S}_{-[d]})$  is sufficient for  $\phi_{[d-k]}$ , for each  $k$ , then the conditional distribution of  $S_k$  given  $(\mathbf{S}_{[k-1]} = \mathbf{s}_{[k-1]}, \mathbf{S}_{-[d]} = \mathbf{s}_{-[d]})$  does not depend on  $\phi_{[d-k]}$  and the FD (5) becomes the product of the “marginal” FDs of the  $\phi_i$ ’s. As a consequence these latter are independent under  $h_s$  which no longer depends on the inferential importance ordering of the parameters.

Even in the multiparameter case it is possible to establish a strict connection between FDs and objective Bayesian posteriors. In this case fiducial priors are related to reference priors: one of the best choices for a default analysis. Notice that reference priors, as FDs, depend on the inferential importance of the parameters.

### 3 Asymptotics for fiducial and/or confidence distributions: the real parameter case

#### 3.1 An expansion with error of order $O(n^{-1})$

In Veronese & Melilli (2015) the asymptotic normality of the FD and/or the CD for the mean parameter of a real NEF is proved and an expansion with an error of order  $O(n^{-1})$  is derived. Here we generalize these results to an arbitrary regular model indexed by a real parameter  $\theta$ . In the following we will only refer to an FD, but the results also hold for a CD.

Let  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$  be an i.i.d. sample of size  $n$  from the density (with respect to the Lebesgue measure)  $p_\theta$ , with  $\theta$  belonging to an open set  $\Theta \subseteq \mathbb{R}$  and let  $\hat{\theta}$  be the maximum likelihood estimator (MLE) of  $\theta$  based on  $\mathbf{X}$ . Let  $\ell(\theta) = n^{-1} \sum_{i=1}^n \log p_\theta(\mathbf{X}_i)$  and let  $\ell''(\hat{\theta})$  and  $\ell'''(\hat{\theta})$  be the second and the third derivative of  $\ell(\theta)$  with respect to  $\theta$ , evaluated in  $\hat{\theta}$ . Then the observed Fisher information per unit is given by  $-\ell''(\hat{\theta})$ , while the expected one is  $I(\theta) = -E_\theta(\partial^2/\partial \theta^2 \log p_\theta(\mathbf{X}_i))$ . Let  $b = b(\hat{\theta}) = -1/\ell''(\hat{\theta})$ . Consider now  $Z = (n/b)^{1/2}(\theta - \hat{\theta})$ , which is an approximate standardized version of  $\theta$  in the fiducial setup, and let  $H_{n,\hat{\theta}}(z)$  be its FD derived from the sampling distribution of  $\hat{\theta}$ . Notice that if  $\hat{\theta}$  is sufficient,  $H_{n,\hat{\theta}}(z)$  is the exact FD for  $\theta$ , otherwise it represents a natural approximation of the exact one, see e.g. Schweder & Hjort (2016). To prove our result we use

the expansion of the frequentist probability  $\Pr_\theta(Z \leq z)$  provided in Datta & Ghosh (1995) or in Mukerjee & Ghosh (1997). Thus we assume that  $p_\theta$  satisfies the regularity assumptions required in the previous papers, see also Ghosh (1994, ch. 8) for further details.

**Theorem 1** *Let  $\mathbf{X}$  be an i.i.d. sample of size  $n$  from a density  $p_\theta$ ,  $\theta \in \Theta \subseteq \mathbb{R}$ . Then, under regularity assumptions, the fiducial distribution  $H_{n,\hat{\theta}}(z)$  of  $Z = (n/b)^{1/2}(\theta - \hat{\theta})$  can be expanded as*

$$H_{n,\hat{\theta}}(z) = \Phi(z) - \phi(z) \left[ \frac{1}{6} b^{3/2} \ell'''(\hat{\theta})(z^2 - 1) \right] n^{-1/2} + O(n^{-1}). \quad (7)$$

If  $p_\theta$  also satisfies the conditions required by Johnson (1970), the following result holds.

**Corollary 1** *In the same setting of Theorem 1, let  $\pi^J(\theta) \propto I(\theta)^{1/2}$  be the Jeffreys prior for  $\theta$ . If  $\pi^J$  is improper, assume that there exists an  $n_0 \geq 1$  such that the posterior distribution of  $\theta$  is proper for  $n \geq n_0$ , almost surely for all  $\theta$ . Then the expansion of  $\Pi^J(\theta|z)$  coincides with that of  $H_{n,\hat{\theta}}(z)$  given in (7).*

Theorem 1 and Corollary 1 naturally establish a connection between FDs and *matching priors*, i.e. priors that ensure approximate frequentist validity of posterior credible sets. For a general review and a specific bibliography on matching priors, see Datta & Mukerjee (2004). For a regular model indexed by a real parameter it is well known, see Datta & Mukerjee (2004, Theorem 2.5.1), that the Jeffreys prior  $\pi^J$  is the unique first order matching prior, i.e. a prior  $\pi$  for which  $\Pr_\theta(\theta \leq q_{1-\alpha}(\mathbf{X}, \pi)) = 1 - \alpha + O(n^{-1})$ , where  $q_{1-\alpha}(\mathbf{X}, \pi)$  denotes the  $(1 - \alpha)$ th posterior quantile of  $\theta$ . Furthermore,  $\pi^J$  is also a second order matching prior if and only if the model satisfies the following condition

$$I(\theta)^{-3/2} E_\theta[(\partial/\partial\theta \ell(\theta))^3] \quad \text{is a constant free of } \theta. \quad (8)$$

Now, because the FD, being also a CD, realizes the exact matching, we have immediately the following

**Corollary 2** *If a fiducial prior  $\pi^F$  exists, then it coincides with the Jeffreys prior  $\pi^J$ . Furthermore, the condition (8) is necessary for the existence of  $\pi^F$ .*

Notice that for a model belonging to a NEF, with mean parameter  $\mu$  and variance function  $V(\mu)$ , condition (8) becomes: “ $2V'(\mu)V(\mu)^{-1/2}$  is independent of  $\mu$ ”. The solution of the implied differential equation is  $V(\mu) = (c_1\mu + c_2)^2$ , i.e. the variance function of the model must be quadratic in the mean. This result was found for the first time in Veronese & Melilli (2015), using a totally different approach.

*Example 1* (Fisher’s gamma hyperbola-Nile problem). Fisher (1973, Sec.VI.9) considers a sample of size  $n$  from a curved exponential family obtained by two independent gamma distributions

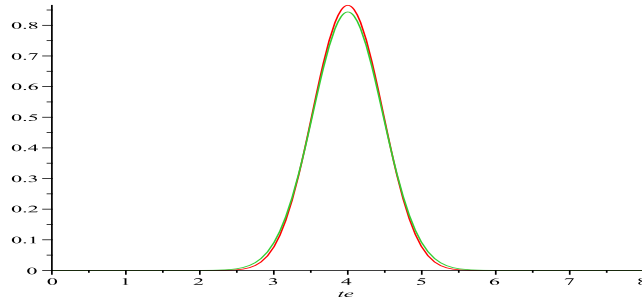


Figure 1: Exact (red) and approximate (green) fiducial densities for  $\eta$  in the Fisher's gamma hyperbola for a sample size  $n = 5$ .

with means constrained on an hyperbole. Following Efron & Hinkley (1978), we directly start with the sufficient statistic  $\mathbf{S} = (S_1, S_2)$ , with  $S_1$  and  $S_2$  distributed according to  $\text{ga}(n, e^{-\eta})$  and  $\text{ga}(n, e^{\eta})$ , respectively. Here  $\text{ga}(\alpha, \beta)$  denotes a gamma distribution with shape parameter  $\alpha$  and mean  $\alpha/\beta$ . Then the likelihood of the model can be written as  $L_{\eta}(\mathbf{s}) = \exp\{-e^{-\eta}s_1 - e^{\eta}s_2\}$ . The MLE  $\hat{\eta} = (1/2)\log(S_1/S_2)$  of  $\eta$  is clearly not sufficient and thus an exact inference requires the use of an ancillary statistic as specified by Fisher (1973), see also Efron & Hinkley (1978) and Barndorff-Nielsen (1980). Here we compute an asymptotic FD for  $\eta$  starting directly from  $\hat{\eta}$ , but notice that the definition of FD given in (5) involves an ancillary statistic; we will come back on this point in the next section. Because it is easy to check that  $E_{\eta}[(\partial\ell(\eta)/\partial\eta)^3] = 0$ , condition (8) holds and thus by Corollary 2 a fiducial prior might exist for this model. Furthermore, since  $\ell'''(\hat{\eta}) = 0$ , it follows from (7) that the normal distribution  $N(\hat{\eta}, b/n)$ , where  $b = -1/\ell''(\hat{\eta}) = n/(2\sqrt{s_1 s_2})$ , represents an approximate FD of  $\eta$  with error of order  $O(n^{-1})$ . Figure 1 reports the asymptotic FD for  $\eta$  compared with the exact one which will be derived in the next section. It shows how excellent is the approximation even for a very small sample size ( $n = 5$  in the plot).

Another type of prior matching studied in the literature directly involves cumulative distribution functions. Because  $H_{n,\hat{\theta}}(z)$  is stochastic in a frequentist setup, the matching is made up between  $E_{\theta}(H_{n,\hat{\theta}}(z)) = E_{\theta}\left(\Pr_{\hat{\theta}}\left\{\sqrt{n/b}(\theta - \hat{\theta}) \leq z\right\}\right)$  and  $\Pr_{\theta}\left\{\sqrt{n/b}(\theta - \hat{\theta}) \leq z\right\}$ , see Datta & Mukerjee (2004, sec. 3.2). Clearly quantiles and distribution functions are strongly connected and thus it is not surprising that the conditions for the existence of matching priors in the two approaches are related. Indeed, the first order matching conditions are the same, while this is not true for the second order ones. Notice that the matching is obtained using the quantity  $Z$  which is an approximate standardization of  $\theta$ . This is meaningful in an asymptotic setting, but it is not appropriate for small sample sizes. In this case the FD realizes an exact matching if we consider instead of  $Z$  the natural

pivotal quantity given by the distribution function of  $\hat{\theta}$ , namely  $F_{\hat{\theta}}(\hat{\theta})$ . We have

$$\begin{aligned} E_{\theta} \left( \Pr_{\hat{\theta}} \{ F_{\hat{\theta}}(\hat{\theta}) \leq z \} \right) &= E_{\theta} \left( \Pr_{\hat{\theta}} \{ 1 - H_{\hat{\theta}}(\theta) \leq z \} \right) = \\ E_{\theta} \left( \Pr_{\hat{\theta}} \{ H_{\hat{\theta}}(\theta) \geq 1 - z \} \right) &= 1 - E_{\theta} \left( \Pr_{\hat{\theta}} \{ H_{\hat{\theta}}(\theta) \leq 1 - z \} \right) = 1 - E_{\theta}(1 - z) = z, \end{aligned}$$

and because  $\Pr_{\hat{\theta}} \{ F_{\hat{\theta}}(\hat{\theta}) \leq z \} = z$ , the exact matching for distribution functions holds. However, exact FDs do not always exist and thus it is natural to look for approximations which have nice asymptotic properties. As a consequence the comparison of distribution functions in terms of the standardized quantity  $Z$  becomes relevant. Furthermore, in a multiparameter case quantiles are not well defined and thus the study of the frequentist properties of a multivariate FD can be conducted along the lines developed for matching distribution functions.

### 3.2 An approximation based on the Barndorff-Nielsen $p^*$ -formula

Consider a sample  $\mathbf{X}$  (or a sufficient statistic  $\mathbf{S}$ ) whose distribution depends on a real parameter  $\theta$ . In the previous section we have obtained an approximate FD for  $\theta$  starting from the distribution of the MLE  $\hat{\theta}$ . However, if  $\hat{\theta}$  is not sufficient, the approximation of the FD can be improved if one also consider the remaining information included in the sample. This can be done using (5) after having applied the “conditionality resolution” to the sampling distribution, i.e. the construction of an ancillary statistic  $A$  and of an approximate conditional distribution of  $\hat{\theta}$  given  $A = a$ , see Barndorff-Nielsen (1980, 1983). This approximation is quite simple and is generally accurate to order  $O(n^{-1})$ , or even  $O(n^{-3/2})$ , and exact in specific cases. Here the term approximation refers to one of the two following situations: i) there exists an ancillary statistic  $A$ , but we are not able to construct the exact conditional distribution of  $\hat{\theta}$  given  $A = a$ ; ii) an exact ancillary statistic does not exist and an approximate one is used. The construction of an FD in case i) is not different in essence from the widespread procedure used to derive a Bayesian posterior starting from an approximate (e.g. profile, pseudo or composite) likelihood. A similar approach is used also by Schweder & Hjort (2016) to construct CDs.

We recall here only some useful features of Barndorff-Nielsen (1980, 1983)’s approach and we refer to these papers for a detailed discussion. The approximate distribution of  $\hat{\theta}$  given  $A = a$  is

$$p_{\theta}^*(\hat{\theta}|a) = c(a, \theta) |j(\hat{\theta})|^{1/2} \frac{L(\theta; \mathbf{x})}{L(\hat{\theta}; \mathbf{x})}, \quad (9)$$

where  $L(\theta, \mathbf{x})$  is the likelihood function,  $j(\hat{\theta})$  is the observed Fisher information and  $c(a, \theta)$  is the normalizing constant which does not depend on  $\theta$  in many important cases. It is worthy to remark that formula (9) is invariant to reparameterizations and is exact for transformation models. Furthermore, under repeated sampling for a real NEF, where no conditioning is involved,  $p^*$  in (9)

is often of order  $O(n^{-3/2})$  and is exact for normal (known variance), gamma (known shape) and inverse-gaussian (known shape) distributions.

If  $F_\theta^*(\hat{\theta}|a)$  denotes the distribution function corresponding to  $p_\theta^*(\hat{\theta}|a)$ , and satisfies the usual conditions reported after (4), we can construct an approximate FD for  $\theta$  as  $h_\theta^*(\theta) = |\partial F_\theta^*(\hat{\theta}|a)/\partial \theta|$ . A first useful result concerns a real NEF for which the exact distribution of  $\hat{\theta}$ , which is clearly sufficient, is difficult to obtain.

**Proposition 2** *Given an i.i.d. sample from a real regular NEF, with density  $p_\theta(x) = \exp\{\theta x - M(\theta)\}$ , then  $h_\theta^*(\theta) = |\partial F_\theta^*(\hat{\theta})/\partial \theta|$  is an approximate fiducial density for  $\theta$ . The order of approximation depends on that of  $p^*$ .*

The following examples, concerning curved exponential families, i.e. NEFs in which a constraint on the natural parameter space is imposed, illustrate a typical case in which formula (9) can be fruitfully applied to construct an FD.

*Example 1 ctd.* As previously observed, the MLE  $\hat{\eta}$  is not sufficient and thus the exact FD can be obtained starting from the conditional distribution of  $\hat{\eta}$  given an ancillary statistic  $A$ . Using  $A = \sqrt{S_1 S_2}/n$ , after some calculations, one obtains

$$p_\eta(\hat{\eta}|a) = \exp\{-2na \cosh(\hat{\eta} - \eta)\} / (2K_0(2na)), \quad (10)$$

where  $K_0(w) = \int_0^\infty \exp\{-w \cosh(z)\} dz$  is the modified Bessel function of the second order evaluated in  $(0, w)$ . As observed by Efron & Hinkley (1978), from (10) it follows that this example involves a (not immediately obvious) translation model, and thus a transformation model, so that  $p_\eta(\hat{\eta}|a) = p_\eta^*(\hat{\eta}|a)$  by Barndorff-Nielsen (1983, Theorem 1). Thus an exact FD for  $\eta$  is  $h_{\hat{\eta},a}(\eta) = -\partial F_\eta^*(\hat{\eta}|a)/\partial \eta$  and, because  $\eta$  is a location parameter, it coincides with the Jeffreys posterior, see Veronese & Melilli (2016, Prop. 8).

*Example 2* (Bivariate normal model). Consider an i.i.d. sample  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , from a bivariate normal distribution with expectations 0, variances 1 and correlation coefficient  $\rho$ . This is a curved exponential model with sufficient statistics  $S_1 = \sum_{i=1}^n (X_i^2 + Y_i^2)/2$  and  $S_2 = \sum_{i=1}^n X_i Y_i$ . Despite its simplicity the estimate of  $\rho$  is a challenging problem as shown in Fosdick & Raftery (2012). Both Efron & Hinkley (1978) and Barndorff-Nielsen (1980) use this example to discuss the construction of an approximate ancillary statistic in a conditional inference setting. Their proposals essentially coincide and lead to consider the “affine” ancillary  $A = (S_1 - n)/\sqrt{n(1 + \hat{\rho}^2)}$ , where  $\hat{\rho}$  is the MLE of  $\rho$ .

To discuss the performance of  $h^*$  obtained starting from  $p^*$ , we compare it with other possible asymptotic FDs and with the Bayesian posterior obtained from the Jeffreys prior  $\pi^J(\rho) \propto (\rho^2 + 1)^{1/2}/(1 - \rho^2)$ . In particular, we consider the following FDs:  $h^r$  and  $h^{rstab}$  obtained starting



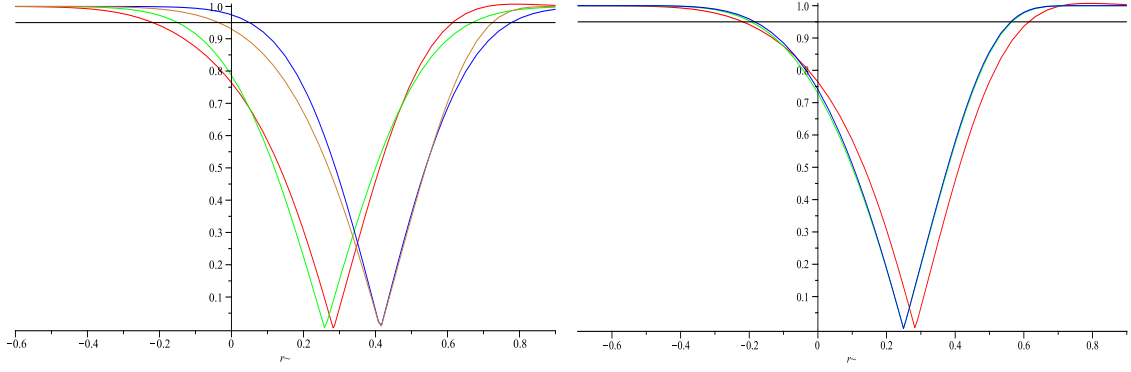


Figure 2: Confidence curves for a sample size  $n = 15$ , generated from  $\rho = 0.3$  with  $s_1 = 19.248$  and  $s_2 = 4.827$ ,  $r = 0.414$  and  $\hat{\rho} = 0.260$ . Left graph:  $cc^r$  (blue),  $cc^{rstab}$  (gold),  $cc^0$  (green) and  $cc^1$  (red). Right graph:  $cc^1$  (red),  $cc^*$  (green),  $cc^J$  (blue). The horizontal line identifies the 95% confidence intervals.

from the approximate normal distribution of the sample correlation coefficient  $r$  and its stabilizing transformation respectively, see Schweder & Hjort (2016, Example 7.2) ;  $h^0$  and  $h^1$  obtained considering the first one or the first two terms of (7), respectively. We assume a sample size  $n = 15$  because a larger value of  $n$ , e.g. 50, produces essentially the same (good) results for all choices. The left graph of Figure 2 reports the confidence curves  $cc^r$ ,  $cc^{rstab}$ ,  $cc^0$  and  $cc^1$  corresponding to the previous FDs. The curves present different behaviors because the two estimators  $r$  and  $\hat{\rho}$  on which they are based assume quite different values in the sample. The right graph compares  $cc^1$  with  $cc^*$  and  $cc^J$  obtained from  $h^*$  and  $\pi^J$ , respectively. As expected, the last two curves, both based on the sufficient statistics  $S_1$  and  $S_2$ , are very similar and induce confidence intervals narrower than those induced by  $cc^1$ . To better appreciate the good behavior of  $h^*$ , we compare the corresponding coverage with those of the other FDs and that of the Jeffreys posterior. Table 1 reports the root mean squared errors multiplied by 1000 for the coverage of the mentioned cases for a sample of size 20 based on 1000 simulated data sets for  $\rho = 0, \pm 0.1, \pm 0.2, \dots, \pm 0.9, \pm 0.95$ . The results confirm the good performance of  $h^*$  which is comparable with  $h^{rstab}$  and  $\pi^J$ .

$h^r$	$h^{rstab}$	$h^0$	$h^1$	$h^*$	$\pi^J$
52.063	6.450	63.469	23.016	7.319	8.087

Table 1: Root mean squared errors multiplied by 1000 for the coverage of the various distributions (1000 simulated samples of size  $n = 20$ , with  $\rho = 0, \pm 0.1, \pm 0.2, \dots, \pm 0.9, \pm 0.95$ ).

Finally, consider a sufficient statistic  $\mathbf{S} = (S_1, S_2)$  with density  $p_{\theta_1, \theta_2}(s_1, s_2) = p_{\theta_2}(s_2|s_1)p_{\theta_1}(s_1)$  for which (9) is exact for both  $p_{\theta_1}(s_1)$  and  $p_{\theta_2}(s_2|s_1)$  for each  $s_1$ . From Barndorff-Nielsen (1983) it follows that, if the normalizing constant in  $p_{\theta_2}^*(s_2|s_1)$  does not depend on  $s_1$ , the  $p^*$  approximation of  $p_{\theta_1, \theta_2}(s_1, s_2)$  is exact. Thus also the FD based on  $p_{\theta_1, \theta_2}^*$  will be exact. It would be interesting to evaluate the performance of the FDs originated in this way when the  $p^*$ -formula is not exact for

one or both distributions.

## 4 Asymptotics for fiducial distributions: the multidimensional parameter case

In Veronese & Melilli (2015) it is shown that the FD for a real NEF is asymptotically normal. Because the multivariate FDs derived in (5) and (6) is a product of one-dimensional conditional FDs, it is quite natural to expect that also the FD for a  $d$ -dimensional NEF is asymptotically normal. This conjecture is strengthened recalling the relationships between fiducial and objective bayesian analysis and the results on the asymptotic normality of posterior distributions. Indeed, the following theorem holds.

**Theorem 2** *Let  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$  be an i.i.d. sample from a regular NEF on  $\mathbb{R}^d$  with  $\mathbf{X}_i$  having density  $p_{\boldsymbol{\theta}}(\mathbf{x}_i) = \exp\{\sum_{k=1}^d \theta_k x_k - M(\boldsymbol{\theta})\}$ , mean vector  $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})$  and variance function  $\mathbf{V}(\boldsymbol{\mu}) = \text{Var}_{\boldsymbol{\mu}}(\mathbf{X}_i)$ . Furthermore, let  $\bar{\mathbf{x}}$  be the observed value of the sample mean  $\bar{\mathbf{X}} = n^{-1} \sum_{i=1}^n \mathbf{X}_i$ . If  $\mathbf{X}_i$  admits bounded density with respect to the Lebesgue measure or is supported by a lattice, then the fiducial distribution of  $\boldsymbol{\mu}$  is asymptotically order-invariant and asymptotically  $N(\bar{\mathbf{x}}, \mathbf{V}(\bar{\mathbf{x}})/n)$ .*

Since  $\mathbf{V}(\bar{\mathbf{x}})$  coincides with the reciprocal of both the observed and the estimated expected Fisher information matrix, recalling standard results about asymptotic posterior distributions, see e.g. Ghosh et al. (2006, Sec. 4.1.2), the following corollary can be stated.

**Corollary 3** *The asymptotically normal fiducial distribution for  $\boldsymbol{\mu}$  obtained in Theorem 2 coincides with the asymptotic Bayesian posterior for  $\boldsymbol{\mu}$ .*

We remark that the proof of Theorem 2 given in the Appendix is different from the standard ones proving asymptotic normality in frequentist or Bayesian settings because it is based on the convergence of conditional distributions.

*Example 3.* Consider a sample of size  $n$  from a multinomial experiment with outcome probability vector  $\mathbf{p} = (p_1, \dots, p_d)$ , with  $\sum_{k=1}^d p_k \leq 1$ . Then the vector of counts  $\mathbf{X} = (X_1, \dots, X_d)$ , with  $\sum_{k=1}^d X_k \leq n$ , is distributed according to a multinomial distribution with parameters  $n$  and  $\mathbf{p}$ . From Theorem 2 it follows that the asymptotic distribution of  $\mathbf{p}$  is normal with mean  $\bar{\mathbf{x}}$  and variance matrix  $\mathbf{V}(\bar{\mathbf{x}})/n$ , where the elements of  $\mathbf{V}(\bar{\mathbf{x}})$  are  $v_{kk} = \bar{x}_k(1 - \bar{x}_k)$  and  $v_{kr} = -\bar{x}_k\bar{x}_r$ ,  $k \neq r$ .

Notice that using the step-by-step procedure described in Section 2, Veronese & Melilli (2016) have derived the exact fiducial density for  $\mathbf{p}$

$$h_{\bar{\mathbf{x}}}^G(\mathbf{p}) \propto \prod_{k=1}^d p_k^{x_k-1/2} \left(1 - \sum_{j=1}^k p_j\right)^{\gamma_k}, \quad \sum_{k=1}^d p_k \leq 1, \quad 0 < p_k < 1,$$

where  $\gamma_k = -1/2$  for  $k = 1, \dots, d-1$  and  $\gamma_d = n - 1/2 - \sum_{j=1}^d x_j$ . This is a generalized Dirichlet distribution and clearly refers to the order of inferential importance, namely  $p_1, \dots, p_d$ . The ordering is of course no longer relevant in the normal asymptotic distribution derived above.

## 5 Appendix

*Proof of Proposition 1.* Veronese & Melilli (2015) provide the two-term Edgeworth expansions for  $H_s$  and  $H_s^\ell$  given by

$$\Phi\left(\frac{z}{\sqrt{V(\bar{x}_n)}}\right) - \phi\left(\frac{z}{\sqrt{V(\bar{x}_n)}}\right) \left[ \frac{V'(\bar{x}_n)}{V(\bar{x}_n)^{3/2}} \frac{2z^2 + V(\bar{x}_n)}{6} \pm \frac{r}{2\sqrt{V(\bar{x}_n)}} \right] n^{-1/2} + O(n^{-1}),$$

where  $r$  is the maximum span of the lattice supporting the NEF, while “+” and “-” identify the expressions for  $H_s$  and  $H_s^\ell$ , respectively. The expansions of the corresponding densities  $h_s$  and  $h_s^\ell$  follow immediately. Recall that  $h_s^G = k_s^{-1} \sqrt{h_s h_s^\ell}$ , where  $k_s = \int \sqrt{h_s(z) h_s^\ell(z)} dz$  is the normalizing constant. We have

$$\begin{aligned} h_s(z) h_s^\ell(z) &= \phi^2\left(\frac{z}{\sqrt{V(\bar{x}_n)}}\right) \frac{1}{V(\bar{x}_n)} \\ &- \phi\left(\frac{z}{\sqrt{V(\bar{x}_n)}}\right) \frac{1}{\sqrt{V(\bar{x}_n)}} \left\{ 2\phi'\left(\frac{z}{\sqrt{V(\bar{x}_n)}}\right) \frac{1}{\sqrt{V(\bar{x}_n)}} \frac{V'(\bar{x}_n)}{V(\bar{x}_n)^{3/2}} \right. \\ &- \left. 2\phi\left(\frac{z}{\sqrt{V(\bar{x}_n)}}\right) \left( \frac{4zV'(\bar{x}_n)}{6V(\bar{x}_n)^{3/2}} \frac{2z^2 + V(\bar{x}_n)}{6} \right) \right\} n^{-1/2} + O(n^{-1}), \end{aligned}$$

so that

$$\begin{aligned} (h_s(z) h_s^\ell(z))^{1/2} &= \phi\left(\frac{z}{\sqrt{V(\bar{x}_n)}}\right) \frac{1}{\sqrt{V(\bar{x}_n)}} - \left\{ \frac{V'(\bar{x}_n)}{V(\bar{x}_n)^2} \frac{2z^2 + V(\bar{x}_n)}{6} \phi'\left(\frac{z}{\sqrt{V(\bar{x}_n)}}\right) \right. \\ &- \left. \frac{2zV'(\bar{x}_n)}{3V(\bar{x}_n)^{3/2}} \phi\left(\frac{z}{\sqrt{V(\bar{x}_n)}}\right) \right\} n^{-1/2} + O(n^{-1}). \end{aligned} \quad (11)$$

Because the expansion of  $k_s$  can be derived as the integral of the previous expansion, using standard properties of the normal density, we have  $k_s = 1 + O(n^{-1})$  and so the expansion of  $h_s^G$  coincides with (11). A simple computation gives the expression (3) for  $H_s^G$  which is equal to the expansion of  $\Pi^J(z|s)$  given by Johnson (1970).  $\diamond$

*Proof of Theorem 1.* For the sake of clearness, in this proof we denote by  $\hat{\Theta}$  the MLE of a parameter  $\theta$  and by  $\hat{\theta}$  the corresponding estimate. Assume that the FD for  $\theta$  can be computed as  $1 - F_\theta(\hat{\theta})$ , where  $F_\theta(\hat{\theta})$  is the distribution function of  $\hat{\Theta}$ , which is assumed decreasing in  $\theta$ . If it is increasing, the FD is  $F_\theta(\hat{\theta})$  and the proof is similar. Then

$$H_{n,\hat{\theta}}(z) = \Pr_{\hat{\theta}} \left\{ \sqrt{n/b}(\theta - \hat{\theta}) \leq z \right\} = \Pr_{\hat{\theta}} \{ \theta \leq \theta_n \} = 1 - \Pr_{\theta_n} \{ \hat{\Theta}_n^* \leq \hat{\theta} \}, \quad (12)$$

where  $\theta_n = z\sqrt{b/n} + \hat{\theta}$  and  $\hat{\Theta}_n^*$  is the MLE based on  $n$  i.i.d. random variables  $X_{n,i}^*, i = 1, \dots, n$ , belonging to the same family of distributions of  $X_i$ , but with parameter  $\theta_n$ . Note that  $\theta_n$  converges to  $\theta$  for  $n \rightarrow \infty$ , because  $\hat{\theta}$  converges to the “true” value  $\theta$  for almost all sequences  $(x_1, x_2, \dots)$  and  $\Theta$  is an open interval. Thus  $\theta_n$  belongs to  $\Theta$  for  $n$  large enough and for each  $z \in \mathbb{R}$ . Starting from (12), we can also write

$$\begin{aligned} H_{n,\hat{\theta}}(z) &= 1 - \Pr_{\theta_n} \{ \sqrt{n/b}(\hat{\Theta}_n^* - \theta_n) \leq \sqrt{n/b}(\hat{\theta} - \theta_n) \} \\ &= \Pr_{\theta_n} \{ \sqrt{n/b}(\hat{\Theta}_n^* - \theta_n) \geq -z \} = \Pr_{\theta_n} \{ \sqrt{n/b}(\theta_n - \hat{\Theta}_n^*) \leq z \}. \end{aligned}$$

Thus, the asymptotic expansion of  $H_{n,\hat{\theta}}(z)$  can be derived by expanding the frequentist distribution function of  $\sqrt{n/b}(\theta_n - \hat{\Theta}_n^*)$ . This expansion can be directly obtained by standard results, even if  $\{X_{n,i}^*, i = 1, 2, \dots, n; n = 1, 2, \dots, \}$  is a triangular array because only random variables and a first order approximation are considered, see e.g. García-Soidán (1998) and Petrov (1995, Theorem 5.22). The frequentist expansion of  $Z = \sqrt{n/b}(\theta - \hat{\Theta})$  has been provided in several papers about matching priors. Using formula (3.2.3) in Datta & Mukerjee (2004) with  $\theta = \theta_n$  and recalling that  $\hat{\Theta}_n^*$  is the MLE of  $\theta_n$ , we obtain

$$\Pr_{\theta_n} \{ \sqrt{n/b}(\theta_n - \hat{\Theta}_n^*) \leq z \} = \Phi(z) - \phi(z) \left[ \frac{1}{2} \frac{I'(\theta_n)}{I(\theta_n)^{3/2}} + \frac{1}{6} \frac{\ell'''(\theta_n)}{(-\ell''(\theta_n))^{3/2}} (z^2 + 2) \right] n^{-1/2} + O(n^{-1}). \quad (13)$$

Now, because  $-\ell''(\hat{\theta}) - I(\hat{\theta}) = O(n^{-1/2})$  (see e.g. Datta & Ghosh, 1995) and  $\theta_n - \hat{\theta} = z\sqrt{b/n} = O(n^{-1/2})$ , we have  $I(\theta_n) = -\ell''(\hat{\theta}) + O(n^{-1/2})$ . Recalling that  $b = -1/\ell''(\hat{\theta})$ , (13) becomes

$$\begin{aligned} \Pr_{\theta_n} \{ \sqrt{n/b}(\theta_n - \hat{\Theta}_n^*) \leq z \} &= \Phi(z) - \phi(z) \left[ -\frac{1}{2} b^{3/2} \ell'''(\hat{\theta}) + \frac{1}{6} b^{3/2} \ell'''(\hat{\theta}) (z^2 + 2) \right] n^{-1/2} + O(n^{-1}) \\ &= \Phi(z) - \phi(z) \left[ \frac{1}{6} b^{3/2} \ell'''(\hat{\theta}) (z^2 - 1) \right] n^{-1/2} + O(n^{-1}), \end{aligned}$$

which is the expansion of  $H_{n,\hat{\theta}}(z)$  stated in the theorem.  $\diamond$

*Proof of Corollary 1.* The result follows immediately using the expansion of the posterior distribution provided by Johnson (1970, Theorem 2.1 and formulae (2.25) and (2.26)), assuming  $\pi^J$  as prior. Notice that under the stated conditions on the posterior, this result can be used even if the prior is improper, as observed in Ghosh et al. (2006, pag. 106).  $\diamond$

*Proof of Proposition 2.* Recalling that for a real NEF  $\bar{x} = M'(\hat{\theta})$ , we can write

$$p_{\theta}^*(\hat{\theta}) = \exp\{n(\theta M'(\hat{\theta}) - M^*(\theta))\},$$

where  $M^*(\theta) = \log(\int \exp\{n(\theta M'(\hat{\theta}))\} d\nu(\hat{\theta}))$ , with  $\nu(\hat{\theta})$  denoting the dominating measure of the density of  $\hat{\theta}$ . Thus  $p_{\theta}^*(\hat{\theta})$  belongs to a regular real NEF and the result follows immediately by Veronese & Melilli (2015, Theorem 1).  $\diamond$

*Proof of Theorem 2.* Given a square  $d \times d$  matrix  $\mathbf{A}$ , we use  $\mathbf{A}_{k[r]}$  and  $\mathbf{A}_{[k][k]}$  to denote the vector of the first  $r$  elements of the  $k$ -th row of  $\mathbf{A}$  and the matrix identified by the first  $k$  rows and columns of  $\mathbf{A}$ . Moreover,  $\mathbf{A}^T$  denotes the transpose of  $\mathbf{A}$ .

In order to determine the asymptotic FD of  $\boldsymbol{\mu}$  we apply the step-by-step procedure introduced in Section 2 to the conditional distribution of  $\bar{X}_k$  given  $\bar{\mathbf{X}}_{[k-1]} = \bar{\mathbf{x}}_{[k-1]}$  for each  $k$ . Clearly for  $k = 1$ , we have the marginal distribution of  $\bar{X}_1$ . Since the covariance matrix  $\mathbf{V}(\boldsymbol{\mu})$  of  $\mathbf{X}_i$  is finite, by the central limit theorem  $\bar{\mathbf{X}}$  is asymptotically  $N_d(\boldsymbol{\mu}, n^{-1}\mathbf{V}(\bar{\mathbf{x}}))$  and thus the marginal distribution of  $\bar{\mathbf{X}}_{[k]}$  is also asymptotically normal with  $E(\bar{\mathbf{X}}_{[k]}) = \boldsymbol{\mu}_{[k]}$  and  $Var(\bar{\mathbf{X}}_{[k]}) = n^{-1}\mathbf{V}(\bar{\mathbf{x}})_{[k][k]}$ . Let

$$\lambda_k = \mu_k + \mathbf{V}(\bar{\mathbf{x}})_{k[k-1]} [\mathbf{V}(\bar{\mathbf{x}})_{[k-1][k-1]}]^{-1} (\bar{\mathbf{x}}_{[k-1]} - \boldsymbol{\mu}_{[k-1]}) \quad (14)$$

and

$$q_k = \mathbf{V}(\bar{\mathbf{x}})_{kk} - \mathbf{V}(\bar{\mathbf{x}})_{k[k-1]} [\mathbf{V}(\bar{\mathbf{x}})_{[k-1][k-1]}]^{-1} [\mathbf{V}(\bar{\mathbf{x}})_{k[k-1]}]^T. \quad (15)$$

Using known results about the convergence of conditional distributions, see Steck (1957, Theorem 2.4) or Barndorff-Nielsen & Cox (1979, Sec.4), under the assumptions of the present theorem it follows that the conditional distribution of  $\bar{X}_k$  given  $\bar{\mathbf{X}}_{[k-1]} = \bar{\mathbf{x}}_{[k-1]}$  is asymptotically  $N(\lambda_k, n^{-1}q_k)$ .

Now recall that for a NEF it is always possible to consider the so called “mixed parameterization”  $(\boldsymbol{\mu}_{[k]}, \boldsymbol{\theta}_{-[k]})$  which is one-to-one with the natural parameter  $\boldsymbol{\theta}$ , see e.g. Brown (1986, ch. 3). For  $\boldsymbol{\theta}_{-[k]}$  fixed, the distribution of  $\bar{\mathbf{X}}_{[k]}$  belongs to a NEF with parameter  $\boldsymbol{\theta}_{[k]}$  and thus the conditional distribution of  $\bar{X}_k$  given  $\bar{\mathbf{X}}_{[k-1]} = \bar{\mathbf{x}}_{[k-1]}$  depends only on  $\theta_k$ . The same must be true of course for the corresponding asymptotic distribution, so that its mean parameter  $\lambda_k$  depends only on  $\theta_k$  and hence only on  $\mu_k$ . Considering now the alternative mixed parameter  $(\boldsymbol{\mu}_{[k-1]}, \theta_k, \boldsymbol{\theta}_{-[k]})$ , it follows that there exists a one-to-one correspondence between  $\theta_k$  and  $\mu_k$ , for  $\boldsymbol{\mu}_{[k-1]}$  and  $\boldsymbol{\theta}_{-[k]}$  fixed. As a consequence  $\boldsymbol{\mu}_{[k-1]}$  can be fixed arbitrarily in the mixed parameterizations  $(\boldsymbol{\mu}_{[k-1]}, \theta_k, \boldsymbol{\theta}_{-[k]})$  and we specifically assume  $\boldsymbol{\mu}_{[k-1]} = \bar{\mathbf{x}}_{[k-1]}$ . Using the parameter  $(\bar{\mathbf{x}}_{[k-1]}, \mu_k, \boldsymbol{\theta}_{-[k]})$ , we have that  $\lambda_k$  coincides with  $\mu_k$ , see (14). Summing up, each of the three parameters  $\lambda_k$ ,  $\theta_k$  and  $\mu_k$  represents a possible parameterization of the asymptotic conditional distribution of  $\bar{X}_k$  given  $\bar{\mathbf{X}}_{[k-1]} = \bar{\mathbf{x}}_{[k-1]}$ , for fixed  $\bar{\mathbf{x}}_{[k-1]}$  and  $\boldsymbol{\theta}_{-[k]}$ . Thus we can find the asymptotic FD of  $\lambda_k$ . Consider now a random vector  $\bar{\mathbf{X}}^*$  with distribution belonging to the same family of that of  $\bar{\mathbf{X}}$ , with mixed parameter  $(\bar{\mathbf{x}}_{[k-1]}, \mu_k^*, \boldsymbol{\theta}_{-[k]})$ , where  $\mu_k^* = \bar{x}_k + z_k/\sqrt{n}$ , with  $z_k \in \mathbb{R}$ , as in the proof of Theorem 1. Notice that the marginal distributions of  $\bar{\mathbf{X}}_{[k-1]}^*$  and of  $\bar{\mathbf{X}}_{[k-1]}$  are equal. Such a  $\mu_k^*$  is well defined for large  $n$  since  $(\bar{\mathbf{x}}_{[k-1]}, \bar{x}_k, \boldsymbol{\theta}_{-[k]})$  is a possible value for the mixed parameter in the distribution of the whole vector, because the NEF is regular and thus the parameter space is open.

For  $n$  varying and fixed  $k$ , the sequence of marginal sample means  $\bar{\mathbf{X}}_{[k]}^*$  derives from random vectors whose mean parameter depends on  $n$ , so that it forms a triangular array. In order to determine the FD of  $\lambda_k$ , we can consider the quantity  $\sqrt{n}(\lambda_k - \bar{x}_k)$ , which is a sort of standardization

of  $\lambda_k$  in our fiducial context. Using (2), similarly to what done in (12), we can write

$$\begin{aligned} \Pr_{\bar{x}_k} \left( \sqrt{n}(\lambda_k - \bar{x}_k) \leq z_k | \bar{\mathbf{X}}_{[k-1]}^* = \bar{\mathbf{x}}_{[k-1]}, \boldsymbol{\theta}_{-[k]} \right) &= \Pr_{\bar{x}_k} \left( \lambda_k \leq \bar{x}_k + \frac{z_k}{\sqrt{n}} | \bar{\mathbf{X}}_{[k-1]}^* = \bar{\mathbf{x}}_{[k-1]}, \boldsymbol{\theta}_{-[k]} \right) \\ &= 1 - \Pr_{\lambda_k^*} \left( \bar{X}_k^* \leq \bar{x}_k | \bar{\mathbf{X}}_{[k-1]}^* = \bar{\mathbf{x}}_{[k-1]}, \boldsymbol{\theta}_{-[k]} \right), \end{aligned} \quad (16)$$

where  $\lambda_k^* = \bar{x}_k + z_k/\sqrt{n}$ . Since  $\text{Var}(\bar{\mathbf{X}}_{[k]}^*)$  is a continuous function of  $\boldsymbol{\mu}^* = E(\bar{\mathbf{X}}^*)$ , it converges to a positive definite matrix for each  $k$  when  $\boldsymbol{\mu}^*$  converges to the “true” value of  $\boldsymbol{\mu}$ , for  $n \rightarrow \infty$ . Then, using the result on the convergence of a conditional distribution presented at the beginning of the proof with  $\boldsymbol{\mu}$  replaced by  $\boldsymbol{\mu}^*$ , we have that  $\bar{X}_k^*$  given  $\bar{\mathbf{X}}_{[k-1]}^* = \bar{\mathbf{x}}_{[k-1]}$  is asymptotically  $N(\lambda_k^*, q_k/n)$ . Notice that from the existence of the second moment of each component of  $\bar{\mathbf{X}}_{[k-1]}^*$ , it follows that the condition required by Steck (1957, Theorem 2.4, formula (28)), for the case of triangular arrays, is satisfied. Thus, the asymptotic normality of  $\bar{X}_k^*$  given  $\bar{\mathbf{X}}_{[k-1]}^* = \bar{\mathbf{x}}_{[k-1]}$  implies, for  $n \rightarrow +\infty$ ,

$$\sup_{z_k} \left| \Pr_{\lambda_k^*} \left( \bar{X}_k^* \leq \bar{x}_k | \bar{\mathbf{X}}_{[k-1]}^* = \bar{\mathbf{x}}_{[k-1]}, \boldsymbol{\theta}_{-[k]} \right) - \Phi \left( \sqrt{\frac{n}{q_k}} (\bar{x}_k - \lambda_k^*) \right) \right| \rightarrow 0 \quad a.s.$$

Recalling the expression of  $\lambda_k^*$ , we obtain

$$\sup_{z_k} \left| \Pr_{\bar{x}_k + z_k/\sqrt{n}} \left( \bar{X}_k \leq \bar{x}_k | \bar{\mathbf{X}}_{[k-1]} = \bar{\mathbf{x}}_{[k-1]}, \boldsymbol{\theta}_{-[k]} \right) - \Phi \left( -z_k/\sqrt{q_k} \right) \right| \rightarrow 0 \quad a.s.$$

which, using (16), gives

$$\sup_{z_k} \left| \Pr_{\bar{x}_k} \left( \sqrt{n}(\lambda_k - \bar{x}_k) \leq z_k | \bar{\mathbf{X}}_{[k-1]} = \bar{\mathbf{x}}_{[k-1]}, \boldsymbol{\theta}_{-[k]} \right) - \Phi \left( z_k/\sqrt{q_k} \right) \right| \rightarrow 0 \quad a.s.$$

We can conclude that the conditional FD of  $\lambda_k$  given  $\boldsymbol{\theta}_{-[k]}$  is asymptotically normal with mean  $\bar{x}_k$  and variance  $n^{-1}q_k$ , and thus it does not depend on  $\boldsymbol{\theta}_{-[k]}$ . Recalling the one-to-one correspondence between  $\theta_k$  and  $\lambda_k$ , for fixed  $\boldsymbol{\theta}_{-[k]}$ , and in particular that  $\lambda_d$  is a one-to-one function of  $\theta_d$ , it follows that  $\lambda_1, \lambda_2, \dots, \lambda_d$  are asymptotically independent, so that the full vector  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_d)$  is asymptotically  $N(\bar{\mathbf{x}}, n^{-1}\mathbf{Q}(\bar{\mathbf{x}}))$ , where  $\mathbf{Q}(\bar{\mathbf{x}})$  is the diagonal matrix with  $k$ -th element  $q_k$ .

To obtain the asymptotic FD of  $\boldsymbol{\mu}$  we consider the one-to-one lower-triangular transformation  $\boldsymbol{\mu} = \mathbf{g}(\boldsymbol{\lambda})$ , with  $\boldsymbol{\lambda} = \mathbf{g}^{-1}(\boldsymbol{\mu})$  given by (14) for  $k = 1, \dots, d$ . Consider now the lower  $d \times d$  triangular matrix  $\mathbf{A} = \mathbf{A}(\bar{\mathbf{x}})$  whose  $k$ -th row is made up by the vector  $-\mathbf{V}(\bar{\mathbf{x}})_{k[k-1]}[\mathbf{V}(\bar{\mathbf{x}})_{[k-1][k-1]}]^{-1}$ , in the first  $k-1$  positions, 1 in the  $k$ -th position and zero elsewhere. Thus we can write  $\boldsymbol{\lambda} = \mathbf{A}\boldsymbol{\mu} + (\mathbf{I} - \mathbf{A})\bar{\mathbf{x}}$  and  $\boldsymbol{\mu} = \mathbf{A}^{-1}\boldsymbol{\lambda} + (\mathbf{I} - \mathbf{A}^{-1})\bar{\mathbf{x}}$ , with  $\mathbf{I}$  denoting the identity matrix of order  $d$ . By applying the Cramér delta method it follows that  $\boldsymbol{\mu}$  is asymptotically normal with (asymptotic) mean and covariance matrix  $\mathbf{A}^{-1}\bar{\mathbf{x}} + (\mathbf{I} - \mathbf{A}^{-1})\bar{\mathbf{x}} = \bar{\mathbf{x}}$  and  $n^{-1}\mathbf{A}^{-1}\mathbf{Q}(\bar{\mathbf{x}})\mathbf{A}^{-1\text{T}}$ , respectively. We now show that  $\mathbf{A}^{-1}\mathbf{Q}(\bar{\mathbf{x}})\mathbf{A}^{-1\text{T}} = \mathbf{V}(\bar{\mathbf{x}})$  or, equivalently,  $\mathbf{Q}(\bar{\mathbf{x}}) = \mathbf{A}\mathbf{V}(\bar{\mathbf{x}})\mathbf{A}^{\text{T}}$ . By direct computation it is easy to see that the  $(k, h)$ -th element of  $\mathbf{A}\mathbf{V}(\bar{\mathbf{x}})$ ,  $k, h = 1, 2, \dots, d$ , is

$$\mathbf{V}(\bar{\mathbf{x}})_{kh} - \mathbf{V}(\bar{\mathbf{x}})_{k[k-1]}[\mathbf{V}(\bar{\mathbf{x}})_{[k-1][k-1]}]^{-1}\mathbf{V}(\bar{\mathbf{x}})_{h[k-1]}^{\text{T}}. \quad (17)$$

Notice that (17) is 0 for  $k > h$  because the product of its last two factors gives a  $(k-1)$ -dimensional vector with 1 in the  $h$ -th position and zero otherwise. The matrix  $\mathbf{AV}(\bar{\mathbf{x}})\mathbf{A}^T$  is of course symmetric, so that it is sufficient to proceed only for  $k \geq h$ . On its diagonal we have

$$\mathbf{V}(\bar{\mathbf{x}})_{kk} - \mathbf{V}(\bar{\mathbf{x}})_{k[k-1]}[\mathbf{V}(\bar{\mathbf{x}})_{[k-1][k-1]}]^{-1}\mathbf{V}(\bar{\mathbf{x}})_{k[k-1]}^T, \quad k = 1, \dots, d, \quad (18)$$

because the only nonzero element in the product of the  $k$ -th row of  $\mathbf{AV}(\bar{\mathbf{x}})$  and the  $k$ -th column of  $\mathbf{A}^T$  is the product of (17), with  $h = k$ , and 1. For  $k > h$ , the  $(k, h)$ -th element of  $\mathbf{AV}(\bar{\mathbf{x}})\mathbf{A}^T$  is zero, because the first  $k-1$  components of the  $k$ -th row of  $\mathbf{AV}(\bar{\mathbf{x}})$  and the last  $d-h$  components of the  $h$ -th column of  $\mathbf{A}^T$  are zero. Thus the matrix  $\mathbf{AV}(\bar{\mathbf{x}})\mathbf{A}^T$  coincides with  $\mathbf{Q}(\bar{\mathbf{x}})$  and this completes the proof of the theorem.  $\diamond$

## 6 Acknowledgments

This research was supported by grants from Bocconi University.

## References

- Barndorff-Nielsen, O. (1980). Conditionality resolutions. *Biometrika*, 67, 293–310.
- Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, 70, 343–365.
- Barndorff-Nielsen, O. & Cox, D. R. (1979). Edgeworth and saddle-point approximation with statistical applications. *J. R. Stat. Soc. Ser. B*, 41, 279–312.
- Brown, L. D. (1986). Fundamentals of statistical exponential families with applications in statistical decision theory. *Lecture Notes-Monograph Series*, 9, 1–279.
- Datta, G. S. & Ghosh, J. K. (1995). On priors proving frequentist validity for Bayesian inference. *Biometrika*, 82, 37–45.
- Datta, G. S. & Mukerjee, R. (2004). Probability matching priors: higher order asymptotics. *Lecture Notes in Statistics*, 178, 1–126.
- Efron, B. & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, 65, 457–482.
- Fisher, R. A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society*, 26, 528–535.

- Fisher, R. A. (1973). *Statistical methods and scientific inference*. Hafner Press: New York.
- Fosdick, B. K. & Raftery, A. E. (2012). Estimating the correlation in bivariate normal data with known variances and small sample size. *The American Statistician*, 66, 34–41.
- García-Soidán, P. H. (1998). Edgeworth expansions for triangular arrays. *Communications in Statistics-Theory and Methods*, 27(3), 705–722.
- Ghosh, J. K. (1994). *Higher order asymptotic*. Institute of Mathematical Statistics and American Statistical Association: Hayward, California.
- Ghosh, J. K., Delampady, M., & Samanta, T. (2006). *An introduction to Bayesian analysis*. Springer: New York.
- Hannig, J., Iyer, H. K., Lai, R. C. S., & Lee, T. C. M. (2016). Generalized fiducial inference: A review and new results. *J. American Statist. Assoc.*, 44, 476–483.
- Johnson, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist*, 41, 851–864.
- Lindley, D. V. (1958). Fiducial distributions and Bayes theorem. *J. R. Stat. Soc. Ser. B*, 20, 102–107.
- Mukerjee, R. & Ghosh, M. (1997). Second-order probability matching priors. *Biometrika*, 84, 970–975.
- Petrov, V. V. (1995). *Limit theorems of probability theory*. Clarendon Press: Oxford.
- Schweder, T. & Hjort, N. L. (2016). *Confidence, likelihood and probability*. London: Cambridge University Press.
- Steck, G. P. (1957). Limit theorems for conditional distributions. *Univ. California Publ. Statist*, 2, 237–284.
- Veronese, P. & Melilli, E. (2015). Fiducial and confidence distributions for real exponential families. *Scand. J. Stat.*, 42, 471–484.
- Veronese, P. & Melilli, E. (2016). Objective Bayesian and fiducial inference: some results and comparisons. *arXiv:1612.01882 [math.ST]*, (pp. 1–37).